The EMBO Journal vol.5 no.13 pp.3625-3630, 1986

Phylogenetic evidence for the acquisition of ribosomal RNA introns subsequent to the divergence of some of the major *Tetrahymena* groups

Mitchell L.Sogin, Amie Ingold, Mark Karlok, Henrik Nielsen[1] and Jan Engberg[1]

Department of Molecular and Cellular Biology, National Jewish Hospital and Research Center, Denver, CO 80206, USA, and [1]Biochemical Institute B, Panum Institute, University of Copenhagen, Copenhagen, Denmark

Communicated by H.Klenow

Previous work has demonstrated the presence of a self-splicing intron in the the large subunit ribosomal RNA coding region in some strains of the ciliate protozoan *Tetrahymena*. Sequence comparisons of the intron regions from six *Tetrahymena* species showed these to fall into three homology groups. In an attempt to evaluate the evolutionary origins of the intervening sequences, we have now determined complete small subunit ribosomal RNA gene sequences from 13 species of *Tetrahymena* and the absolute number of nucleotide differences between the sequences was used to construct a phylogenetic tree. This phylogeny was consistent with the groupings suggested by comparisons of other biochemical characters including cytoskeletal proteins, isozyme analyses, and restriction maps of complete rRNA transcription units. The homology groupings that were based upon the intron sequence data do not agree with the relationships inferred from the small subunit rRNA sequence data. These observations are taken to indicate that the intron character has been acquired independently in different species at a stage later than the branching out of the species.

Key words: small subunit rRNA sequence/intron evolution/phylogenetic trees/*Tetrahymena*

Introduction

Introns in the large subunit ribosomal RNA (rRNA) genes have been found in such diverse eukaryotic organisms as slime molds, insects and ciliates (Cech *et al.*, 1982; Gerbi *et al.*, 1982; Engberg, 1985). In addition, they can be found in chloroplast and fungal mitochondrial genomes – both considered to be descendents of prokaryotic ancestors (Gray and Doolittle, 1982; McCarroll *et al.*, 1983; Yang *et al.*, 1985). The recent finding of an intron in the large subunit rRNA gene of an archebacterium (Kjems and Garrett, 1985) has added interest to the challenging question of the evolutionary origins of rRNA introns. The nucleotide sequence of the rRNA introns from different genera are not phylogenetically conserved but their location appears to be restricted to a few distinct regions within the large subunit rRNAs which are structurally very nearly constant in all prokaryotic and eukaryotic organisms (Gerbi *et al.*, 1982; Engberg, 1985; Kjems and Garrett, 1985).

The rRNA introns in the ciliate genus *Tetrahymena* have been the subject of intense investigation during the past few years (Cech *et al.*, 1982). This was mainly prompted by the surprising finding that the precursor-rRNA of *Tetrahymena*, which contains an intron of the so-called group I (Michel and Dujon, 1983), is capable of undergoing self-splicing in an *in vitro* reaction that is dependent only on the presence of simple salts and a guanosine nucleotide (Cech *et al.*, 1981). Recently, the complete nucleotide sequences of the large subunit rRNA introns from six species of the genus *Tetrahymena* have been reported (Nielsen and Engberg, 1985a). These introns appear to be well conserved in evolution (compared to introns in structural genes) and they fall into three phylogenetic groupings. Evaluating the significance of the intron relationships is complicated by the lack of consensus in constructing evolutionary relationships or trees for the *Tetrahymena*. Most phylogenetic schemes for *Tetrahymena* are based upon comparisons of morphologic (Corliss, 1973) or biochemical characters (Meyer and Nanney, 1986; Nanney *et al.*, 1980; Nielsen *et al.*, 1985; Williams *et al.*, 1984), but there is little consensus regarding which characters are optimal for inferring relationships. In any event, comparisons of biochemical phenotypes or morphologic characters do not provide a quantitative measure of genetic relatedness.

In contrast rRNA sequences have proven to be effective instruments for dissecting phylogenetic relationships (Stackebrandt and Woese, 1981). Ribosomal RNAs are functionally equivalent in all known organisms and they do not appear to undergo lateral transfer between species. The small subunit rRNAs are particularly useful because they are relatively large (1752–1753 nucleotides long in the genus *Tetrahymena*) and hence provide a statistically significant number of variable positions. They offer the additional attribute of being comprised of a mosaic of genetic regions that display different rates of evolutionary change (Sogin *et al.*, 1986). These regions permit the measurement of both close and distant evolutionary distances.

In an attempt to provide a framework for the comparisons of the evolutionary relationships from the *Tetrahymena* rRNA introns, we have determined the complete small subunit rRNA gene sequence from 13 species of *Tetrahymena* including the previously reported small subunit rRNA gene of *T. thermophila* (Spangler and Blackburn, 1985). The resulting phylogenetic tree does not agree with the homology groupings inferred from comparisons of intron data and hence suggests that large subunit rRNA introns invaded the genome subsequent to the divergence of some of the major *Tetrahymena* groups.

Results

Comparisons of the 13 *Tetrahymena* small subunit rRNA sequences reveal a total of 47 positions that display variation. All the sequences are 1752 nucleotides in length with the exception of *T. thermophila* and *T. malaccensis* which contain an extra cytosine at position 487. The sequences fall into three major homology groups and representative sequences from each of the groups are displayed in Figure 1. Those positions which display variation in any of the examined *Tetrahymena* small subunit rRNA genes are enclosed in boxes and their sequences can be inferred from the data provided in Table I.

Table II shows the homology values and the absolute number of nucleotide differences in pairwise comparisons of the *Tetra-*

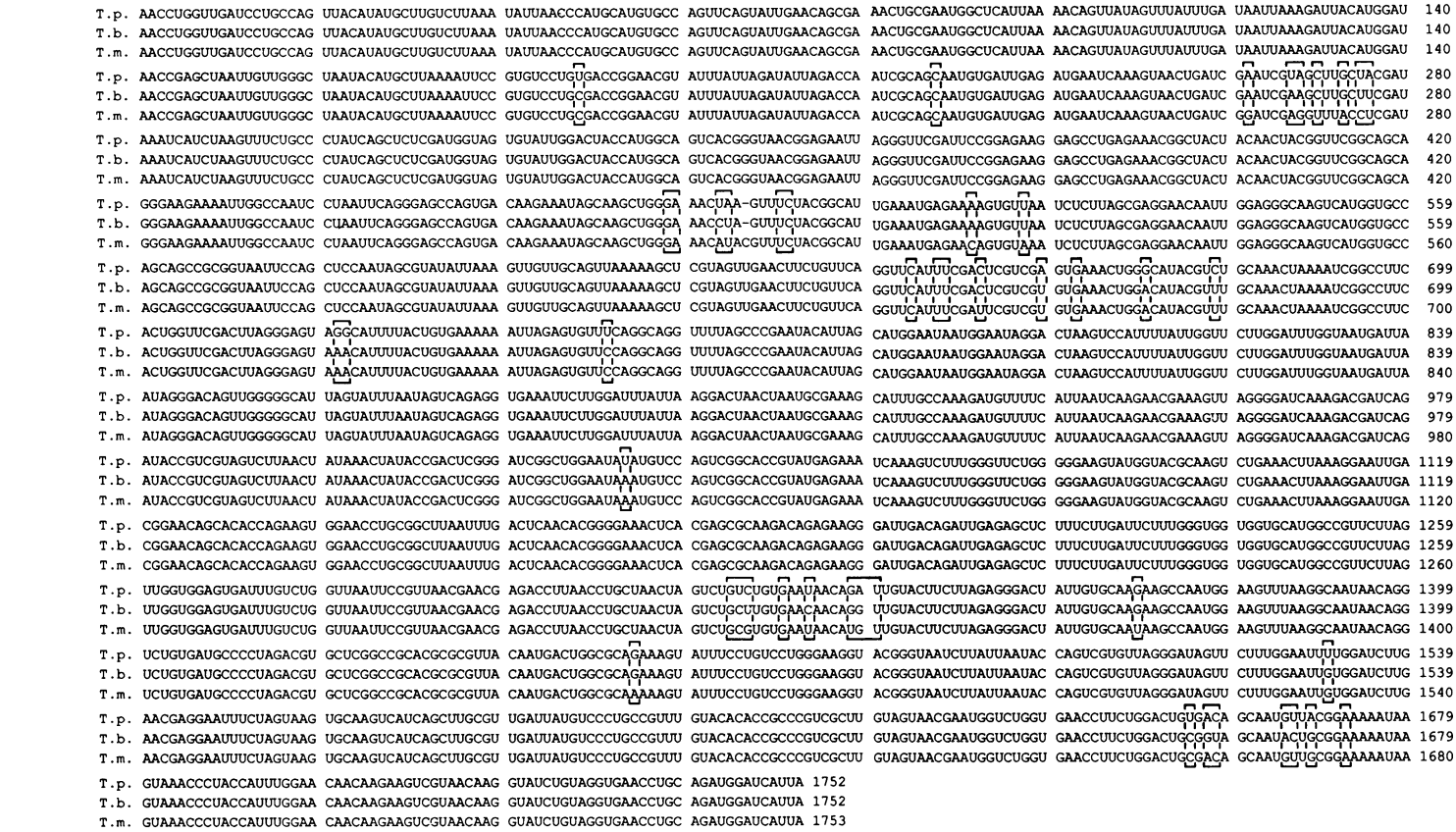

Fig. 1. Alignment of three small subunit rRNA sequences from the genus *Tetrahymena*. Comparisons of small subunit rRNAS from 13 species of *Tetrahymena* demonstrate that they fall into three homology groups with a total of 47 positions displaying variation. The sequences from *T. pigmentosa* (T.p.), *T. borealis* (T.b.) and *T. malaccensis* (T.m.) represent the three groups and positions that are known to vary in at least one of the 13 *Tetrahymena* sequences are enclosed in boxes. A number system for each sequence is provided in the figure.

Table I. Location of sequence variations in three classes of *Tetrahymena* small subunit rRNA genes

T. pigmentosa, T. nanneyi, T. hyperangularis-like rRNA sequences

Organism	Position[d]										
	189	228	479	480	491	492	1325	1326	1334	1340	1341
T. hegewischi[a]	C	–	A	C	G	U	A	C	C	G	–
T. australis[a]	–	U	–	–	–	–	–	C	C	G	C
T. capricornis[a]	C	–	–	–	–	–	–	–	C	–	–
T. patula[a]	–	–	–	–	–	–	–	–	C	–	–

T. borealis, T. canadensis-like rRNA sequences

	Position[d]											
	189	268	484	518	645	648	649	660	663	1658	1667	1673
T. tropicalis[b]	–	G	–	A	U	A	A	–	A	A	U	–
T. pyriformis[b]	U	–	U	A	–	–	–	A	–	–	–	G

T. malaccensis-like rRNA sequences

	Position[d]			
	484	1327	1331	1339
T. thermophila[c]	U	U	A	G

[a]The small subunit rRNA sequences of *T. pigmentosa, T. nanneyi* and *T. hyperangularis* are identical and the positions which vary in the closely related small subunit rRNAs from *T. hegewischi, T. australis, T. capricornis* and *T. patula* are given.
[b]The small subunit rRNA sequences of *T. borealis* and *T. canadensis* are identical and positions which vary in the closely related small subunit rRNAs from *T. tropicalis* and *T. pyriformis* are given.
[c]Positions which vary between *T. malaccensis* and *T. thermophila*.
[d]The numbers refer to the aligned *T. thermophila* sequence positions in Figure 1.

Table II. Homology and nucleotide distance data between *Tetrahymena* small subunit rRNA gene sequences[a]

Organism	*T. heg*	*T. aus*	*T. cap*	*T. pat*	*T. pig*	*T. hyp*	*T. nan*	*T. tro*	*T. pyr*	*T. bor*	*T. can*	*T. mal*	*T. the*
T. hegewischi		0.995	0.996	0.995	0.995	0.995	0.995	0.984	0.986	0.986	0.986	0.981	0.981
T. australis	8		0.997	0.998	0.997	0.997	0.997	0.985	0.988	0.987	0.987	0.983	0.983
T. capricornis	7	5		0.999	0.999	0.999	0.999	0.986	0.987	0.988	0.988	0.983	0.983
T. patula	8	4	1		0.999	0.999	0.999	0.985	0.988	0.987	0.987	0.983	0.983
T. pigmentosa	9	5	2	1		1.000	1.000	0.985	0.987	0.987	0.987	0.983	0.983
T. hyperangularis	9	5	2	1	0		1.000	0.985	0.987	0.987	0.987	0.983	0.983
T. nanneyi	9	5	2	1	0	0		0.985	0.987	0.987	0.987	0.983	0.983
T. tropicalis	28	26	25	26	27	27	27		0.994	0.995	0.995	0.989	0.990
T. pyriformis	25	21	22	21	22	22	22	11		0.997	0.997	0.988	0.989
T. borealis	24	22	21	22	23	23	23	8	5		1.000	0.989	0.990
T. canadensis	24	22	21	22	23	23	23	8	5	0		0.989	0.990
T. malaccensis	33	31	30	31	30	30	30	19	21	19	19		0.998
T. thermophila	33	31	30	31	30	30	30	18	19	18	18	4	

[a]The upper right half of the table gives homology values H for all pair-wise comparisons of *Tetrahymena* small subunit rRNA sequences. We define H as, $H = m/(m+u+g/2)$ where m is the number of sequence positions with matching nucleotides in the two sequences, u is the number of sequence positions with non-matching nucleotides and g is the number of sequence positions that have a gap in one sequence opposite a nucleotide in the other sequence. The absolute number of base changes between nucleotide sequences are shown in the lower half of the table.

hymena small subunit rRNA gene sequences. The extent of sequence variation ranges from no differences in comparisons of *T. nanneyi, T. hyperangularis* and *T. pigmentosa* or in comparisons of *T. borealis* to *T. canadensis*, to as many as 33 differences between *T. hegewischi* and *T. malaccensis*. The absolute number of nucleotide differences between all pairs of *Tetrahymena* species was used in a modification of the distance matrix methods (Fitch and Margoliash, 1967) to construct the phylogenetic tree shown in Figure 2 (Elwood *et al.*, 1985). (A minimum change method generated a tree with an identical topology to that shown in Figure 2, data not shown.) The root of the distance matrix tree was defined by comparing the *Tetrahymena* sequences to more distantly related ciliates including *Oxytricha nova* and *Paramecium tetraurelia*. In this phylogeny the *Tetrahymena* species fall into three major groups and the depth of branching between the groups is similar to the depth of branching between the ciliate genera *Oxytricha* and *Stylonychia* (Elwood *et al.*, 1985).

The evolutionary tree based on the small subunit rRNA sequences provides a framework against which other sets of mutational data can be evaluated. We have recently published a sequence comparison of the rRNA introns from six different species of *Tetrahymena*. The homology values for comparisons of the four intron sequences which display sequence variation are presented in Table III. It can be seen that within the genus *Tetrahymena*, the rRNA introns are considerably more divergent than the small subunit rRNA gene sequences but better conserved than the sequences in the central rRNA spacer regions (data not shown; Engberg, 1983; Nielsen and Engberg, 1985a). The intron$^+$ species listed in Table III are represented in the tree inferred from comparisons of the small subunit rRNA sequences. The striking observation that can be made from inspection of the homology data presented in Table III is that the homology between *T. thermophila* or *T. malaccensis* intron sequences is substantially lower than the homologies between either organism and *T. pigmentosa* and *T. hyperangularis*. These results contradict the evoutionary relationships defined by the comparisons of the small subunit rRNA genes where *T. thermophila* and *T. malaccensis* are seen to be close relatives.

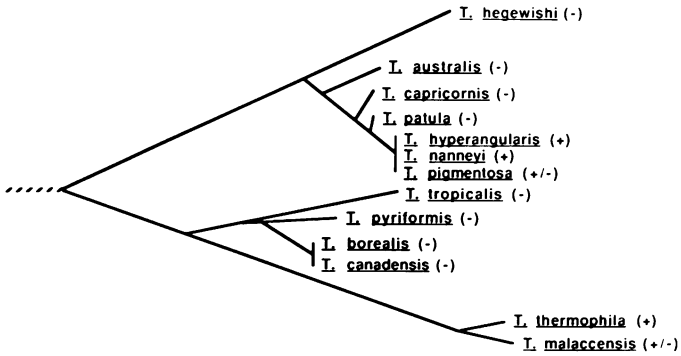

Fig. 2. Molecular phylogenetic tree inferred from *Tetrahymena* small subunit rRNA sequences. A distance matrix tree was inferred by using the absolute number of nucleotide differences in pairwise comparisons of the *Tetrahymena* small subunit rRNA sequences shown in Table II. The tree was constructed using the distance matrix methods of Fitch and Margoliash (1967) as previously described (Elwood *et al.*, 1985). The evolutionary distance between nodes of the tree is represented in the horizontal component of their separation in the figure. The presence (+) or absence (−) of the introns in the large subunit rRNAs of different *Tetrahymena* species is indicated.

Table III. Homology and structural distance data between *Tetrahymena* large subunit rRNA introns[a]

Organism	Structural similarity to			
	T. pig	*T. hyp*	*T. mal*	*T. the*
T. pigmentosa		0.988	0.927	0.932
T. hyperangularis	0.00246		0.930	0.934
T. malaccensis	0.07652	0.07380		0.906
T. thermophila	0.07160	0.06893	0.10072	

Structural distance to (substitutions/site).
[a]The upper right half of the table gives the homology values for aligned large subunit rRNA intron sequences (Nielsen and Engberg, 1985a). The structural distances [average number of base changes per sequence position (Elwood *et al.*, 1985)] are shown in the lower half of the table.

Discussion

Functionally equivalent macromolecules can be used to infer phylogenetic relationships between organisms if the genes defining the molecules share a common ancestry and if they do not undergo lateral transfer between organisms. The validity of a molecular phylogeny can be judged by its agreement with groupings that are based upon other molecular criteria. The branching pattern for the *Tetrahymena* tree depicted in Figure 2 is consistent with the major groupings suggested by comparisons of other

biochemical markers. For example, three clusters can be recognized among pairwise comparisons of cytoskeletal proteins from thirteen *Tetrahymena* species (Williams *et al.*, 1984). In that analysis, *T. borealis, T. canadensis* and *T. tropicalis* form one group while *T. australis, T. capricornis, T. hegewischi, T. hyperangularis* and *T. pigmentosa* form a second cluster. *T. thermophila* appears to be only distantly related to the other two groups. The distribution of restriction endonuclease sites in related gene families has also been studied in *Tetrahymena*. Restriction maps of the ribosomal RNA transcription units for *T. pigmentosa, T. hyperangularis* and *T. nanneyi* are identical, and they are very similar to patterns displayed by *T. capricornis* or *T. australis* and *T. hegewischi*. The rRNA restriction patterns displayed by *T. canadensis* and *T. borealis* are identical and they are closely related to *T. pyriformis*. The *T. thermophila* and *T. malaccensis* patterns are very similar to each other but appear to be only distantly related to the other two major groups. The groupings based upon the rDNA restriction mapping data are consistent with the phylogeny presented in Figure 2.

The homology values, which reflect variation at 47 positions, can be precisely measured for the *Tetrahymena* small subunit rRNA sequences. Although the phylogenetic tree in Figure 2 displays increased resolution relative to the groupings inferred from comparisons with other biochemical phenotypes, the significance of minor branchings must not be over interpreted. The local branching orders for very close relatives, e.g. *T. patula* and *T. capricornis*, cannot be rigorously determined since their small subunit rRNAs differ at a single nucleotide position. In evaluating the small subunit rRNA sequence data it is important to address the question of structural polymorphisms for the rRNA transcription units in strains within different *Tetrahymena* species. Apart from the presence/absence of intervening sequences in some strains of *T. pigmentosa* and *T. malaccensis*, no inter-strain differences could be detected in a restriction enzyme mapping analysis of the rDNA from 11 *Tetrahymena* species (Din and Engberg, 1979). Furthermore, partial sequence analysis of the small subunit rRNA coding regions of the strain C of *T. thermophila* showed complete identity to that of *T. thermophila* strain B (Spangler and Blackburn, 1985). As more biochemical and sequence data from additional sets of evolutionary homologues become available, it will be possible to evaluate the fine structural details of this tree. However, based upon the agreement of the general branching pattern for this tree with groups suggested by other molecular criteria, it appears that the phylogeny in Figure 2 serves as a useful framework for depicting the historical evolution of the organisms considered in our analysis.

In contrast to the small subunit rRNA phylogeny, the homology groupings based upon comparisons of the intron sequence data are not consistent with the phylogenetic relationships suggested by other molecular criteria nor are they consistent with small subunit rRNA phylogenetic tree. The introns are clearly derived from common ancestral sequences but they may be capable of lateral gene transfer between organisms. The contradictory homology groupings displayed by small subunit rRNA sequences with large subunit rRNA intron structures can be explained by proposing that the introns are capable of undergoing lateral gene transfer and invaded the rRNA genes at a time after the separation of some of the major groups of *Tetrahymena*. This notion is further supported by the absence of introns in the large subunit rRNAs of *Tetrahymena* species which diverged before the appearance of intron containing species (see Figure 2). This of course implies multiple independent insertions of the intron in exactly the same position in the *Tetrahymena* large subunit rRNA genes. This is not an unlikely event since all known rRNA introns, even in distantly related species, e.g. *Drosophila* and yeast, are found in the same region of the gene for the large subunit rRNA. As for the rRNA introns in the nuclear gene of *Physarum polycephalum* (a slime mold) and in the mitochondrial gene of yeast, these map at exactly the same position while the rRNA introns of the other species mentioned in the Introduction map at nearby positions (Gerbi *et al.*, 1982; Engberg, 1985). The fact that rRNA introns from different genera do not map at exactly the same position in the large subunit rRNA in itself suggests that the intron character was acquired at different times in evolution; it is unlikely that the ancestral rRNA sequence contained several introns at rather close proximity, and all but one were

Table IV. *Tetrahymena* species, strains and small subunit rRNA clones used in this study

Species	Strain designation	ATCC code	Cloned small subunit rDNA fragment[a,b]	Plasmid name[c]
T. hegewischi	KP7 (intron$^-$)	30832	6.2 kb H−B	pTheg17
T. australis	MGO (intron$^-$)	30831	6.2 kb H−B	pTaus17
T. capricornis	AU-F_1-1 (intron$^-$)	30291	5.2 kb R−B	pTcap17
T. patula	LFF (intron$^-$, macrostome form)	50064	2.9 kb C−C	pTpat17
T. pigmentosa	UM 1286 (intron$^-$)	–	6 kb R−B	pTpig17
Y. hyperangularis	EN 10I (intron$^+$)	30273	6 kb R−B	pThyp17
T. nanneyi	XQ5 (intron$^+$)	30840	6 kb R−B	pTnan17
T. tropicalis	TC3 (intron$^-$)	30352	4 kb H−B	pTtro17
T. pyriformis	GL-C (intron$^-$, a-micronuclear)	–	10 kb K terminus	pKT
T. borealis	UM 731 (intron$^-$)	–	3.8 kb H−B; 0.4 kb H−H	pTbor17; pTbor17-1
T. canadensis	UM 1215 (intron$^-$)	30368	3.8 kb H−B; 0.4 kb H−H	pTcan17; pTcan17-1
T. malaccensis	MP75 (intron$^+$)	50066	5 kb X−B	pTma117
T. thermophila	A-17682a (intron$^+$)	30377	4.2 kb C−B	pTthe17

[a]The letters H, B, R, X, K and C symbolize *Hin*dIII, *Bam*HI, *Eco*RI, *Xba*I, *Kpn*I and *Cla*I restriction enzyme sites, respectively.
[b]Partial restriction enzyme maps of the extrachromosomal rDNA molecules of the listed species have been published previously (Din and Engberg, 1979; Nielsen *et al.*, 1985). The cloned fragment related unambiguously to the restriction maps except in the case of *T. tropicalis* where the cloned fragment was generated by incomplete digestion of the isolated rDNA.
[c]Plasmid pKT was a generous gift of T.Higashinakagawa.

eliminated within a given line of descent. In the specific case of *Tetrahymena* it is difficult to envisage how a precise elimination of an entire intron region could occur. Such elimination processes are usually believed to proceed via reverse transcription of mature transcripts and subsequent integration into the genome. This mechanism is not likely to operate in *Tetrahymena* since the germ line copy of the rRNA gene resides in the micronucleus which is transcriptionally inactive.

The concept that introns are possibly mobile elements was first discussed on the basis of structural evidence obtained from studies on the rDNA introns of *Drosophila* (Rae *et al.*, 1980; Roiha *et al.*, 1981; Dover and Coen, 1981). The presence of a duplication of a 28S rRNA gene sequence that flanks the insert was demonstrated as well as the presence of intron units elsewhere in the genome, also flanked by the same 28S gene segment. The generation of host sequence duplication upon insertion of an extraneous sequence is a common feature of mobile elements. However, these structural features do not fit the description of the rDNA intron of *Tetrahymena* in its present form. The intron flanking sequences are not tandemly repeated and the intron sequence appears to be unique to the rDNA molecule (Wild and Sommer, 1980). Furthermore, the *Tetrahymena* intron does not destroy the transcriptional activity of the integrated rDNA which is in contrast to what seems to be the case in *Drosophila*. The rDNA introns in dipteran flies may, therefore, belong to a different category compared to that of the 'lower' eukaryotes. The questions why the rDNA intron of *Tetrahymena* is not found elsewhere in the *Tetrahymena* genome and why the introns in insect rDNA are located at a position similar to that of the rDNA introns in 'lower' eukaryotes remain unanswered.

One alternative interpretation of our data is that the *T. malaccensis* and *T. thermophila* intron sequences acquired mutations at a much greater rate than did those of *T. hyperangularis* and *T. pigmentosa*. An accelerated rate of genetic change would result in a lower level of similarity between the *T. malaccensis* and the *T. thermophila* introns. Although this interpretation cannot be rigorously eliminated, we consider it unlikely that the introns have evolved at different rates in different *Tetrahymena* species. Altered rates of change could occur if selective pressures beyond maintaining the capacity to undergo self splicing were imposed upon the introns. However, we have previously shown that no selective advantage could be observed in intron$^+$ relative to intron$^-$ rDNA alleles within the same *Tetrahymena* species (Nielsen and Engberg, 1985b).

Materials and methods

Strains and culture methods

All *Tetrahymena* strains, with the exception of *T. pyriformis* and *T. thermophila*, were kindly provided by Dr Ellen Simon of the University of Illinois at Urbana-Champaign. Species and strain designations are given in Table IV. Cell stocks were kept at room temperature in 1.5% (w/v) proteose-peptone, 0.15% (w/v) yeast extract and maintained by transfer every 2 weeks. Cultures for DNA preparation were grown at 30°C with aeration.

Isolation of rDNA and cloning of small subunit rRNA coding fragments

The macronuclear rDNA was isolated either in the form of snap-back molecules or in its native palindromic form by procedures already published (Din and Engberg, 1979; Nielsen *et al.*, 1985) and subsequently digested with restriction enzymes known to generate fragments encompassing the entire small subunit rRNA coding region. Following gel electrophoretic separation on low-temperature gelling agarose gels, the appropriate restriction fragments were isolated and ligated to properly prepared bacterial plasmid vectors using standard DNA recombinant techniques (Maniatis *et al.*, 1982). The vectors used were pACYC184, pBR322 or pSP62-PL (New England Nuclear) and the bacterial host used for transformation was *Escherichia coli* MC1000.

Dideoxynucleotide sequencing

Subsequent to digestion with *Eco*I or *Sal*I, phenol extraction, and recovery by ethanol precipitation, duplex plasmid DNAs were used as templates in a modification of the dideoxynucleotide sequencing protocols (Messing, 1983). Thirty nanograms of synthetic primer complementary to evolutionarily conserved regions of the rRNA genes (Elwood *et al.*, 1985) were annealed to 5 μg of linear duplex DNA in 12 μl of annealing solution [10 mM Tris−HCl (pH 7.5)] by heating at 95°C for 3 min and quick freezing in a dry ice/ethanol bath. These preparations were then thawed and distributed immediately to each of the five tubes containing 6 μl of reaction mix: 10 mM Tris−HCl (pH 8.0), 10 mM $MgCl_2$, 1 mM dithiothreitol, Klenow fragment of DNA polymerase (250 U/ml), 20 μCi [α-^{35}S]-dATP and the deoxynucleotide triphosphates (dNTPs) plus one dideoxynucleotide triphosphate (Elwood *et al.*, 1985). After incubation for 20 min at 37°C, 2 μl of a non-radioactive chase solution (1 mM in all dNTPs) plus additional Klenow enzyme was added and incubation was continued for 15 min. The reactions were halted by addition of EDTA (pH 7.2) to a final concentration of 7.2 mM and then concentrated by ethanol precipitation at −20°C. The precipitates were washed with 80% ethanol/H_2O, dried under vacuum and then suspended in 10 μl of gel loading buffer (0.1% xylene cyanol/0.1% bromophenol blue in formamide). Two microliters from each reaction tube were fractionated on 6% or 8% polyacrylamide sequencing gels (Sanger and Coulson, 1975). Except for short regions at the 5′ and 3′ termini, both the coding and non-coding DNA strands were sequenced for each of the *Tetrahymena* small subunit rRNA genes. Sequencing reactions primed with oligonucleotides complementary to the coding strand cannot access the initial 25 nucleotide positions in the small subunit rRNA gene. Similarly, the 3′ proximal 45 nucleotides cannot be sequenced by reactions that are 'primed' with oligomers complementary to the non-coding strand. No sequence variations were observed in these regions for the examined *Tetrahymena* small subunit rRNAs. Experimental errors were minimized by redundant sequence determinations from each of the oligonucleotide primers. Artifacts resulting from band compressions on polyacrylamide gels were detected by comparing the sequences of both DNA strands and by using an extra ddG reaction in which deoxyinosine (dI) was substituted for deoxyguanosine (dG). (Band-compression artifacts often result in erroneous sequence interpretations and are probably caused by strong secondary interactions that distort gel sieving patterns or effect premature chain terminations in the sequencing reactions. The stacking interactions of dI are weaker than those of dG and thus the secondary structure stabilizers are altered. The reduced stabilities minimize the band compressions that can be detected by comparing the dI containing reactions with similar reactions containing dG.) We estimate that the error rate for the small subunit rRNA sequences presented in this communication is <0.1% and is lower than the likelihood of multiple mutations at positions which differ between any two *Tetrahymena* small subunit rRNA sequences.

Acknowledgements

We are grateful to Professor D.L.Nanney and Dr E.Simon for supplying the different *Tetrahymena* strains. This work was supported by a National Institutes of Health Grant (GM32964) to M.L.S. and by the Danish Natural Science Research Council.

References

Cech,T.R., Zaug,A.J. and Grabowski,P.J. (1981) *Cell*, **27**, 487−496.
Cech,T.R., Zaug,A.J., Grabowski,P.J. and Brehm,,S.L. (1982) In Busch,H. and Kornblum,L. (eds), *The Cell Nucleus*. Academic Press, New York, Vol. 10, pp. 171−204.
Corliss,J.O. (1973) In Elliot,A.M. (ed.), *Biology of Tetrahymena*. Dowden, Hutchison and Ross, Stroudsburg, PA, pp. 1−55.
Din,N. and Engberg,J. (1979) *J. Mol. Biol.*, **1341**, 555−574.
Dover,G. and Coen,E. (1981) *Nature*, **290**, 731−732.
Elwood,H.J., Olsen,G.J. and Sogin,M.L. (1985) *Mol. Biol. Evol.*, **2**, 399−410.
Engberg,J. (1983) *Nucleic Acids Res.*, **11**, 4939−4945.
Engberg,J. (1985) *Eur. J. Cell Biol.*, **35**, 133−151.
Fitch,W.M. and Margoliash,E. (1967) *Science*, **155**, 279−284.
Gerbi,S.A., Gourse,R.L. and Clark,C.G. (1982) In Busch,H. and Kornblum,L. (eds), *The Cell Nucleus*. Academic Press, New York, Vol. 10, pp. 351−386.
Gray,M.W. and Doolittle,W.F. (1982) *Microbiol. Rev.*, **46**, 1−42.
Kjems,J. and Garrett,R.A. (1985) *Nature*, **318**, 675−677.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) In *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
McCarroll,R., Olsen,G.J., Stahl,Y.D., Woese,C.R. and Sogin,M.L. (1983) *Biochemistry*, **22**, 5858−5868.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20−78.
Meyer,E.B. and Nanney,D.L. (1986) *Isozymes: Current Topics in Biological and Medical Research*, in press.

Michel,F. and Dujon,B. (1983) *EMBO J.*, **2**, 33–38.
Nanney,D.L., Simon,E.M. and Whitt,G.S. (1980) *J. Protozool.*, **27**, 451–459.
Nielsen,H. and Engberg,J. (1985a) *Nucleic Acids Res.*, **13**, 7445–7455.
Nielsen,H. and Engberg,J. (1985b) *Biochim. Biophys. Acta*, **825**, 30–38.
Nielsen,H., Simon,H. and Engberg,J. (1985) *J. Protozool.*, **32**, 480–485.
Rae,P.M.M., Kohorn,B.D. and Wade,P. (1980) *Nucleic Acids Res.*, **8**, 3491–3504.
Roiha,H., Miller,J.R., Woods,L.C. and Glover,D.M. (1981) *Nature*, **290**, 749–753.
Sanger,F. and Coulson,A.R. (1975) *J. Mol. Biol.*, **94**, 441–448.
Sogin,M.L., Elwood,H.J. and Gunderson,J.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1383–1387.
Spangler,E.A. and Blackburn,E.H. (1985) *J. Biol. Chem.*, **260**, 6334–6340.
Stackebrandt,E. and Woese,C.R. (1981) In Carlile,M.J., Collins,J.F. and Moseley,B.E.B. (eds), *Molecular and Cellular Aspects of Microbial Evolution*. Cambridge University Press, Cambridge, pp. 1–31.
Wild,M.A. and Sommer,R. (1980) *Nature*, **283**, 693–694.
Williams,N.E., Buhse,H.E. and Smith,M.G. (1984) *J. Protozool.*, **31**, 313–321.
Yang,D., Oyaizu,Y., Oyaizu,H., Olsen,G.J. and Woese,C.R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4443–4447.

Received on 7 August 1986; revised on 30 September 1986